KARNATAKA STATE AKKAMAHADEVI WOMEN'S UNIVERSITY, VIJAYAPURA

DEPARTMENT OF COMPUTER SCIENCE

Synopsis

"Artificial Intelligence techniques and Applications to Agriculture"

Submitted by

Miss. Syeda Bibi Javeriya
Research Scholar
(Doctor of Philosophy)

Research Guide

Dr. Aziz Makandar
Professor
Department of Computer Science
KSAWU, Vijayapura

2021-22

INDEX

SL No.	Topics	Page No.
1	Introduction	1
2	Literature Survey	1
3	Motivation	3
4	Problem Statement	3
5	Scope	3
6	Objectives	3
7	Methodology	6
8	Outcome	7
9	References	8

ABSTRACT

Agriculture plays a vital role in Indian economy and more than 50% of the india's population depend on agriculture. Farming is the pillar for each Nation's economy due to the huge population of India the growth of crop productivity is the major aspect that requires more attention.The production of crops may depend on weather conditions like rainfall, temperature etc. because of which importance of the farming is gradually falling down.The main reason to this is changes in the climate that effects the parameters.Traditional farming is not precise and not based on any analytical details. To overcome such problems like predicting the crop yield will help the farmers to know amount of crop yield in advance, also about the types of crops to be cultivated during the particular climate,Crop recommendation will help for effective decision making,Crop monitoring is also an important factor need attention beside this it will help the farmers to monitor more than one agricultural land at the same time. The usage of various machine learning techniques will provide the benefits for pre-planning the agricultural activities for effective implementation to maximize yield rates and reduce losses.

Synopsis of the research work submitted to the Department of Computer Science KSAWU,Vijayapura.

1	Title of the Thesis	"Artificial Intelligence techniques and Applications to Agricultural Science"
2	Name of the Research Scholar	Miss.Syed Bibi Javeriya
3	Name and Designation of the Research Supervisor	Dr.Aziz Makandar Professor, Dept.of Computer Science KSAWU,Vijayapura.
4	Ph.D Registration Number	KSAWUV/Ph.D/539/2020-21
5	Date of Registration	23-09-2021

Signature of the Candidate

(MISS.SYEDA BIBI JAVERIYA)

(Signature of the Guide with seal)

(Signature of the Chairman with seal)

1.INTRODUCTION

Agriculture is very crucial to India's Market. Around 60 percent of the total land in the country is used for agriculture to meet the needs of 1.2 billion people, so improving crop production is therefore seen as a significant aspect of agriculture.

Farmers have been utilizing age-old ways to anticipate the best planting date for generations. Typically, they would sow in early June to take advantage of the monsoon season, which ran from June to August. As in the past, yield prediction was calculated by analyzing a farmer's previous experience on a particular crop. The Agricultural yield is primarily dependent on weather conditions pests and planning of harvest operation. Accurate information about the history of crop yield is an important thing for making decisions related to agricultural risk management.However, changing weather patterns have resulted in unpredictable monsoons in the last decade, resulting in low crop yields.

Thus, the sowing date becomes crucial in ensuring that farmers harvest a good crop. And if it fails, it results in a loss because a lot of money is spent on seeds and fertilizer applications. Due to atmospheric circumstances and soil changeability, farmers should use the data to generate new insights to make basic cultivating decisions.

In recent times, it has become inevitable to use technology to create awareness about cultivation. The seasonal climatic conditions are also being changed against the fundamental assets like soil, water and air which lead to insecurity of food. In a scenario, crop yield rate is falling short of meeting the demand consistently and there is a need for a smart system which can solve the problem of decreasing crop yield .Therefore eliminating the problem machine learning algorithms are efficiently used in the agriculture field for various purposes.The ML has the capability to perform different calculations and predictions very effectively with minimum time.

2.LITERATURE SURVEY

Jude Immaculate et al. [2], describes the Machine Learning algorithms in agriculture sector. This computes efficiencies for all the machine learning algorithms which are suitable for crop yield prediction. This also explains how the algorithm works and the algorithm implementation. This paper computes all the algorithm efficiencies which are used. This also explains about how well the machine learning techniques are suitable for prediction of crop yield. Limitation: Not all Algorithms are specified in detail and Ensembling techniques are also not used in this paper.

Anuja Chandgude et al. [3], introduces a system where data is retrieved through the sensors. Further, ANN (Artificial Neural Networks) machine learning algorithm is used for data prediction. The dataset contains parameters like humidity, temperature, wind speed and sunlight. Here, crop growth and diseases are predicted using machine learning. This paper focuses more on monitoring using IOT devices. Limitation: This paper focuses more or monitoring using IOT devices rather than prediction.

Pavan Patil et al. [4], introduces an efficient crop recommendation system using classifiers. This paper compared two ML classification Algorithms "Decision Tree and KNN". Data set consists of soil parameters, climatic parameters, production parameters. This paper computes with machine learning algorithms. This paper compared both algorithms individually but has not used combination. This is a model for recommendation system using classifiers suitable from author's point. Limitation: KNN gives less accuracy compared to other algorithms, Ensembling techniques are not used.

K.Ruth Ramya et al. [5], describes about crop yield prediction using random forest classifier. They considered min temp, max temp, humidity and rainfall as a parameter; the results obtained with 99.7% accuracy. Limitation: Only Limitation about this paper is that there is no ensemble algorithm used which reduces the accuracy comparatively

COMPARATIVE ANALYSIS

Authors	Methodology	Advantage	Limitation	Performance matrics
Kumar et al [8]	SVM	The SVM method has implemented a cascade of two SVM classifiers for achieving the accuracy, specificity and precision metrics	The developed model was not given the proper extensive analysis of the defective outlines such as color, shapes and texture. Hence, it is failed to identify the infected surface on the defective patterns	Accuracy=97.77 % Sensitivity=96.5 5% Precision =99.24%
Tiwari and Shukla [9]	CNN, Modified Convolutional Neural Network (MCNN)	The CNN model was developed which utilized spatial features as input and trained by backpropagation that reduced error of prediction as well.	The developed model reduced the relative error as well as decreased the prediction efficiency of crop yield.	MCNN RMSE value = 1396.4 Relative Error=9.8465
Shastry and Sanjay [10]	H-ANN) Hybridized ANN,	The developed (H-ANN) was used to forecast agricultural data such as air temperature and crop yield estimation. IN H-ANN, the LN algorithm was used to train the ANN	The developed (H-ANN) was used to forecast agricultural data such as air temperature and crop yield estimation. IN H-ANN, the LN algorithm was used to train the ANN	RMSE=4.72
Gopal and Bhargavi [11]	ANN and Multiple Linear Regression (MLR).	The developed model is a combination of backpropagation algorithm with ANN to evaluate the exact crop yield.	The developed model is a combination of backpropagation algorithm with ANN to evaluate the exact crop yield.	MLR RMSE=9.8% MAE=6.9% R=89% ANN RMSE=5.1% MAE=6.4% R=99%

3.MOTIVATION

Agriculture is the most important sector that influences the economy of India. It contributes to 18% of India‘s Gross Domestic Product (GDP) and gives employment to 50% of the population of India. People of India are practicing Agriculture for years but the results are never satisfying due to various factors that affect the crop yield [12]. To fulfill the needs of around 1.2 billion people, it is very important to have a good yield of crops. Due to factors like soil type, precipitation, seed quality, lack of technical facilities etc. the crop yield is directly influenced. To focuses on implementing crop yield prediction system by using Machine learning techniques by doing analysis on agriculture dataset. For evaluating performance Accuracy is used as one of the factors.

4.PROBLEM STATEMENT

The problem statement revolves to predict the crop yield, monitor health of the crops and recommend the soil suitable crops by taking into consideration the most influencing parameters to get a better selection of crops which can be grown over the season this can be solved using various machine learning algorithms. This would help reduce the difficulties faced by the farmers in selecting the crop to get high yield and thus maximize profits which in turn will reduce the suicide rates.

5.SCOPE

Integrating farming and machine learning can lead to further advancements in agriculture by maximizing yield and optimizing the use of resources involved. Previous year‘s production data is an essential element for predicting the current yield. The goal is to help the farmers by combining agriculture and technology.

6.OBJECTIVES

The goal is predicting the crop yield at a particular weather condition and thereby recommending suitable crops for that field and monitoring the crops with the efficient technology.

1. Collect the crop dataset and preprocess the dataset. It is done to remove inaccurate, incomplete and unreasonable data ,that increases the quality of the data.

2. Perform Exploratory Data Analysis (EDA) that helps in analyzing the complete dataset and summarizing the main characteristics. It is used to discover patterns, spot anomalies and to get graphical representations of various attributes. Most importantly, it tells us the importance of each attribute, the dependence of each attribute on the class attribute and other crucial information.

3. Divide the analysed crop data into training and testing sets and train the model using the training data to predict the crop yield , recommend suitable crops and monitor health of the crops for given inputs.

4. Compare various Algorithms by passing the analysed dataset through them and calculating the error rate and accuracy for each. Choose the algorithm with the highest accuracy and lowest error rate.

5. Designing the user friendly application based on the observed predictions.

MACHINE LEARNING

Machine Learning is undeniably one of the most influential and powerful technologies in today‘s world. Machine learning is a tool for turning information into knowledge. In the past 50 years, there has been an explosion of data [6]. This mass of data is useless; we analyse it and find the patterns hidden within. Machine learning techniques are used to automatically find the valuable underlying patterns within complex data that we would otherwise struggle to discover. The hidden patterns and knowledge about a problem can be used to predict future events and perform all kinds of complex decision making. To learn the rules governing a phenomenon, machines have to go through a learning process, trying different rules and learning from how well they perform. Hence, why it‘s known as Machine Learning.

Basic Terminology:

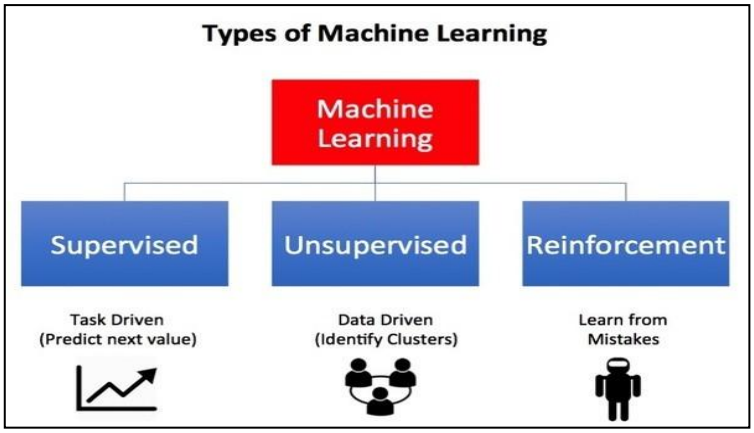
1.Dataset: A set of data examples, which contain features important to solving the problem.

2. Features: Important pieces of data that help us understand a problem. These are fed into a Machine Learning algorithm to help it learn.

3. Model: The representation (internal model) of a phenomenon that a Machine Learning algorithm has learnt. It learns this from the data it is shown during training. The model is the output you get after training an algorithm. For example, a decision tree algorithm would be trained and produce a decision tree model.

Types of Machine Learning

There are multiple forms of Machine Learning; supervised, unsupervised, semisupervised and reinforcement learning. Each form of Machine Learning has differing approaches, but they all follow the same underlying process and theory.



Basic Process

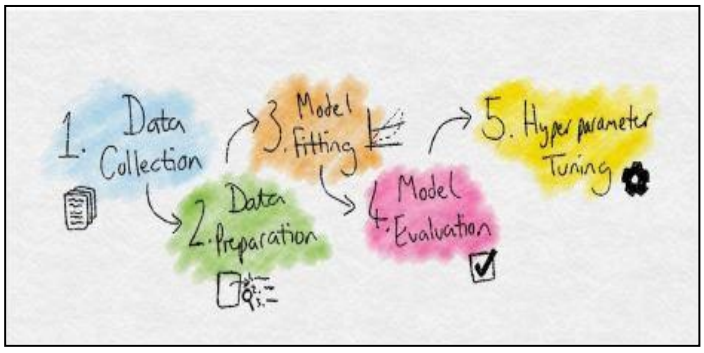
1.Data Collection: Collect the data that the algorithm will learn from.

2.Data Preparation: Format and engineer the data into the optimal format, extracting important features and performing dimensionality reduction.

3.Training: Also known as the fitting stage, this is where the Machine Learning algorithm actually learns by showing it the data that has been collected and prepared.

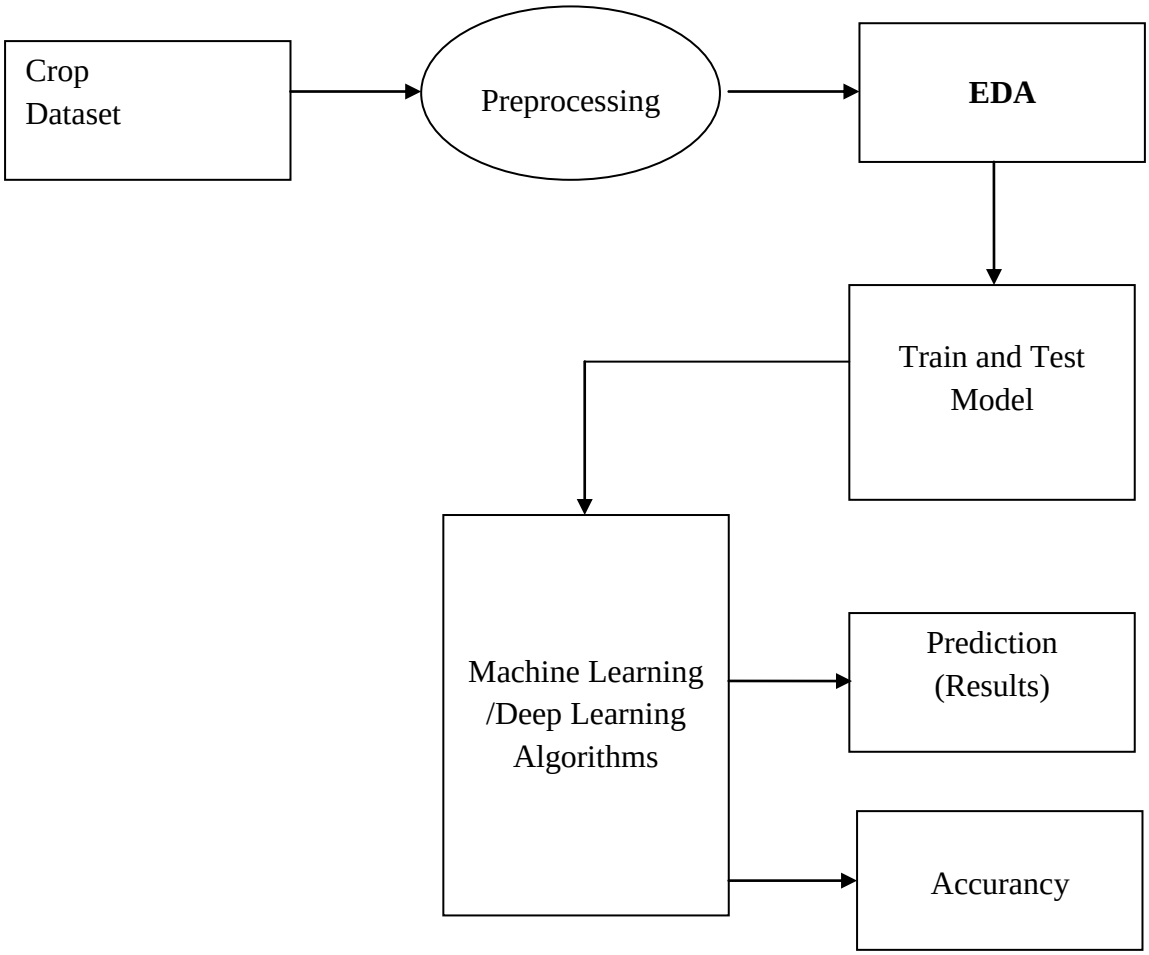
4. Evaluation: Test the model to see how well it performs.

5. Tuning: Fine tune the model to maximize its performance.Fine tune the model to maximize its performance.



General process

7. METHODOLOGY



WORKFLOW

Machine learning uses historical data and information to gain experiences and generate a trained model by training it with the data. The model then makes output predictions. The better the collection of dataset, the better will be the accuracy of the classifier. It has been observed that machine learning methods such as regression and classification perform better than various statistical models[13].

REGRESSION: Regression is a supervised learning technique which helps in finding the correlation between variables and enables us to predict the continuous output variable based on the one or more predictor variables. It is mainly used for prediction, forecasting, time series modeling, and determining the causal-effect relationship between variables.

"Regression shows a line or curve that passes through all the datapoints on target-predictor graph in such a way that the vertical distance between the datapoints and the regression line is minimum."

Some terminologies of regression :

- **Dependent Variable:** The main factor in Regression analysis which we want to predict or understand is called the dependent variable. It is also called target variable.

- **Independent Variable:** The factors which affect the dependent variables or which are used to predict the values of the dependent variables are called independent variable, also called as a predictor.
- **Outliers:** Outlier is an observation which contains either very low value or very high value in comparison to other observed values. An outlier may hamper the result, so it should be avoided.
- **Multicollinearity:** If the independent variables are highly correlated with each other than other variables, then such condition is called Multicollinearity. It should not be present in the dataset, because it creates problem while ranking the most affecting variable.
- **Underfitting and Overfitting:** If our algorithm works well with the training dataset but not well with test dataset, then such problem is called Overfitting. And if our algorithm does not perform well even with training dataset, then such problem is called underfitting.

  By performing the regression, we can confidently determine the most important factor, the least important factor, and how each factor is affecting the other factors[14].

  ClASSIFICATION: The Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations on the basis of training data. In Classification, a program learns from the given dataset or observations and then classifies new observation into a number of classes or groups.

Classification is of two types:

1.Binary Classification: When we have to categorize given data into 2 distinct classes.

2.Multiclass Classification: The number of classes is more than 2.

8. OUTCOME

Estimating Crop yield before the harvest is one of the most significant concern in agriculture since crop yield variations from year to year impact on farmer's agricultural policies.crop recommendation will help the farmers to grow suitable crops on a particular soil and weather conditions ,crop monitoring will help the farmers to take necessary actions accordingly to produce expected crops and prevent from the loss due to pests. The main objective is to find best features for better predictions using various machine learning Algorithms.Crop yield information using ML algorithms helps the farmers to make quick decisions for upcoming situations such as choosing alternative crops and weathers to abondon a crop at early stage of growth.

REFERENCE

[2].Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. Sensors, 18(8), 2674.

[3].Chandgude, A., Harpale, N., Jadhav, D., Pawar, P., & Patil, S. M. (2018). A Review on Machine Learning Algorithm Used For Crop Monitoring System in Agriculture. International Research Journal of Engineering and Technology (IRJET), 5(04), 1470.

[4]. Patil, A., Kokate, S., Patil, P., Panpatil, V., & Sapkal, R. (2020). Crop Prediction using Machine Learning Algorithms. International Journal of Advancements in Engineering & Technology, 1(1), 1-8.

[5].van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture, 177, 105709.

[6]. https://www.agricultureinformation.com/forums/threads/agriculture-in-india.101537/.

[7]. https://towardsdatascience.com/machine-learning-basics-part-1-a36d38c7916.

[8]. S. D. Kumar, S. Esakkirajan, S. Bama, and B. Keerthiveena, “A microcontroller based machine vision approach for tomato grading and sorting using SVM classifier,” Microprocessors and Microsystems, vol. 76, pp.103090, 2020.

[9]. P. Tiwari, and P. Shukla, “Crop yield prediction by modified convolutional neural network and geographical indexes,” International Journal of Computer Sciences and Engineering, vol. 6, no. 8, pp. 503-513, 2018.

[10]. K. A. Shastry, and H. A. Sanjay, “Hybrid prediction strategy to predict agricultural information,” Applied Soft Computing, vol. 98, pp. 106811, 2021.

[11]. P. M. Gopal, and R. Bhargavi, “A novel approach for efficient crop yield prediction,” Computers and Electronics in Agriculture, vol. 165, pp. 104968, 2019.

[12].N. P. Sastra and D. M. Wiharta, ―Environmental monitoring as an IoT application in building smart campus of UniversitasUdayana,‖ in Proc. Int. Conf. Smart Green Technol. Elect. Inf. Syst. (ICSGTEIS), Oct. 2016, pp. 85–88.

[13].https://healthcare.ai/machine-learning-versus-statistics-use/.

[14]. https://www.javatpoint.com/regression-analysis-in-machine-learning.